II. STATISTICAL MACHINE TRANSLATION

II.4 INTRODUCTION

PROBLEM: TRANSLATE SENTENCE IN SOURCE LANGUAGE 1 ("FRENCH") INTO A SENTENCE IN TARGET LANGUAGE 2.

IDEA: USE $\hat{s} = \arg \max_{\hat{s}} [\mathcal{P}(\hat{s} | 1) \mathcal{P}_2(\hat{s})]$ TRANSLATION MODEL

NEW ISSUE: ALIGNMENT

DENOTE ALIGNMENTS BY $a(\hat{s}, \tilde{s})$

$\hat{s}$ HAS LENGTH $l$
$\tilde{s}$ HAS LENGTH $m$

TOTAL NUMBER OF ALL POSSIBLE ALIGNMENTS

$2^{-lm}$

EXAMPLE OF CORRECT ALIGNMENT
11.2 TRANSLATION MODELS

**NOTATION**

\[ d = d_1 = 1, d_i, i, e \text{, from English vocabulary} \]

\[ \bar{d} = d_1, d_2, \ldots, d_m \text{, from French vocabulary} \]

\[ c = c_1 = 1, c_i, \ldots, c_n \]

\[ y = \{ 0, 1 \} \text{ connected} \]

\[ y = 1 \text{ means that the} \quad \text{French word word at position} \ y \text{ is connected to the English word at position} \ i \]

\[ y = 0 \text{ means: not connected to English word} \]

**GENERAL EXACT DECOMPOSITION**

\[ P(d, \bar{d}, \{1\}) = P(m, \{1\}) \prod_{i=1}^{m} P(a_i, 1, d_i, d_i-1, m, i) \]

- Expand conditional probabilities into fractions to proof.
- Need to make independence assumptions.
Assume:

\[ P(c_m | e) = \text{independent of } m \text{ and } e \]

\[ P(c_m | d_1^{m-1}, d_m^2) = \frac{1}{m} \quad \text{uniform} \]

\[ P(c_i | d_1^{i-1}, d_i^2, m, e) = \prod_{j=1}^{m} F(c_i | d_j, e_j) \]

\[ P(c_f | d_i) = \frac{t}{(e + 1)^m} \sum_{d_i = 0}^{e+1} \prod_{j=0}^{m} F(c_j | d_j, e_j) \]

For any \( \alpha \) we need

\[ P(c_f | e) = \frac{t}{(e + 1)^m} \sum_{d_i = 0}^{e+1} \prod_{j=0}^{m} F(c_j | d_j, e_j) \]

Training:

Maximum likelihood with constraint \[ \sum_{i} F(c_i | e) = 1 \quad \text{(i.e. \textit{vocabulary})} \]

\[ L = \frac{t}{(e + 1)^m} \sum_{d_i = 0}^{e+1} \prod_{j=0}^{m} F(c_j | d_j, e_j) \]

\[ - \sum_{i} \left( e F(c_i | e) - 1 \right) \]
\[ E(\theta_1) = \frac{1}{n} \sum_{i=1}^{n} p(\theta_1, \theta_i) \sum_{j=1}^{\text{number of times } x_i \text{ is connected to } j} \delta_{c_i, c_j} \]

- Iterative Scheme
- GM-Style Training

IBM Model 2

Explicitly model alignment by

\[ a(\alpha_i, l, m, e) = P(\alpha_i | \theta_1 \ldots, \theta_i, m, e) \]

with \[ \sum_{i=0}^{\infty} a(\alpha_i | l, m, e) = 1 \]

\[ P(\theta_1, \theta_i) = \frac{1}{\mathcal{Z}} \exp \left( \sum_{i=0}^{\infty} \prod_{j=0}^{m} E(\theta_j | \theta_1, \theta_i) \cdot a(\alpha_i | l, m, e) \right) \]

Training: Maximum Likelihood

Using advance multipliers
IBM MODEL 3-5

- Also model "fictitious":
  How many French words should be connected to an English word

- Make alignment depend on words